

Typing and linking (geographic) data workflows

Simon Scheider,
Universiteit Utrecht,
Department of Human Geography
and Spatial Planning



Workflow tools



Support workflow design, documentation and re-computation:

- Check of data types
- Error checks and execution control
- Wings: export of linked data

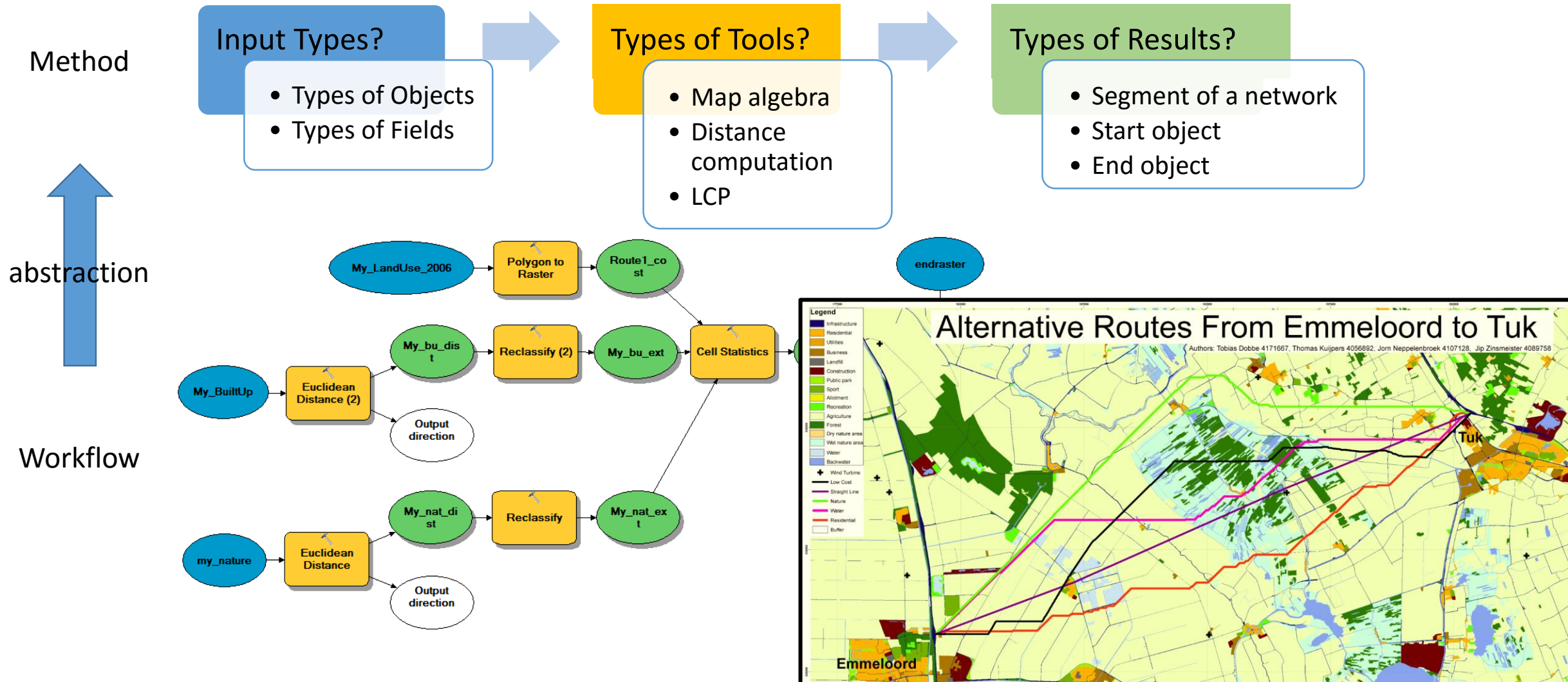
But workflows not easily re-usable as methods:

- Manual annotation
- Not easily published and searched
- Not easily modularized and adapted
- No recommendation

System	Annotation granularity	Annotation languages	Annotation generation	Semantic richness	Recommender system	Annotation abstraction level	External resources
Kepler	Nodes	XML	Manual	Low	None	Natural language	Ontologies, Web Services
Knime	Nodes	XML	Manual	Low	None	Natural language	
Orange	Project		Manual	Low	None	Natural language	
Taverna	Model		Manual	Low	None	Formal and natural language	MyExperiment, Web services
VisTrails	Model	XML	Manual	Low	None	Natural language	
Pegasus	Model	SQL	Both	Low	None	Natural language	Crowdlabs, Web services
WINGS	Model	OWL	Automated	High	Data	Formal and natural language	
ArcGIS model builder	Nodes	Python	Manual	Low	None	Natural language	Ontologies, Web services

What makes a (GIS) workflow reusable?

Semantic types and modules

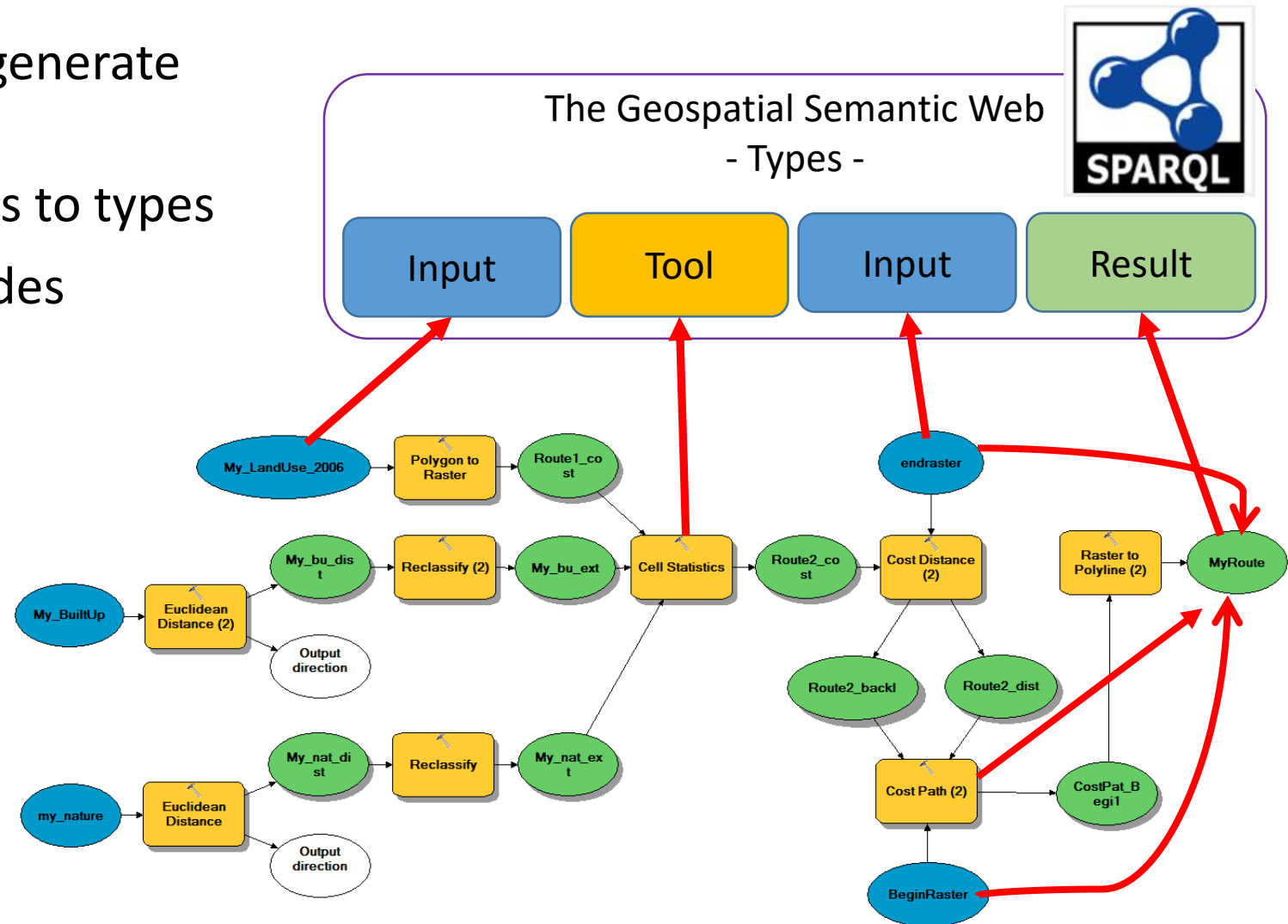


Sharing (GIS) methods on the Web

Capture workflows and generate semantic **links**:

1. From workflow nodes to types
2. Among workflow nodes

Query workflows with SPARQL

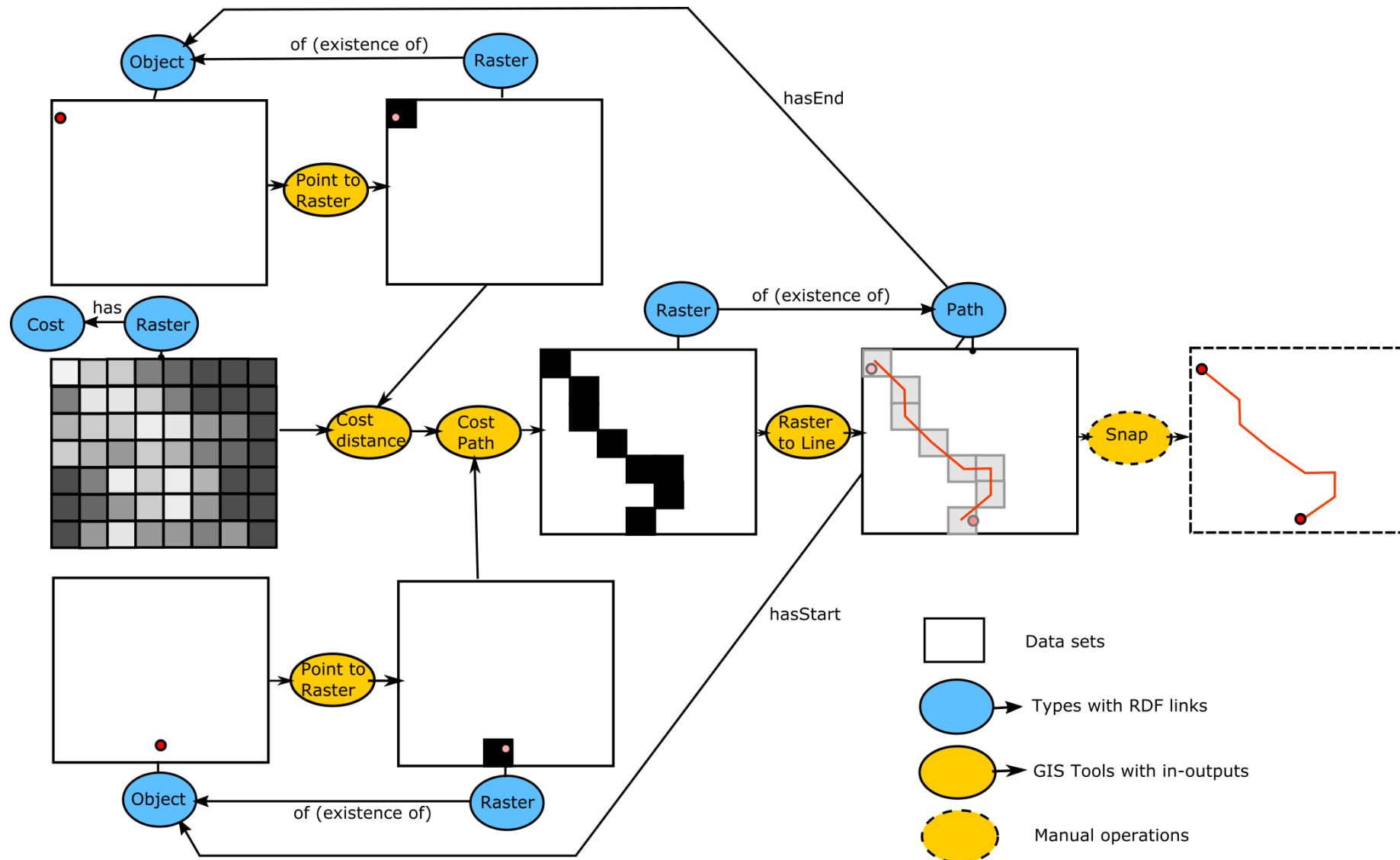


Typing and enrichment principle

Tools are sources of semantic information. They

- ... constrain the semantics of results and inputs over and above data types
- ... link origin data sets with result data sets
- ... can thus be used to enrich and type data as well as workflows

... by example (LCP road design)



What does this buy us?

- ... automatic typing of data results with semantic concepts
(e.g. “network paths, fields, objects, events, trajectories, ...”)
- ... automatic integration of data
(e.g. “snapping paths with start and end”)
- ... search for workflows based on types of results or inputs
(e.g. “methods for generating spatial networks”)
- ... recommendation of meaningful next steps
(e.g. “network integration”)

Challenges

- Typing of tools?
- Type propagation (forward and backward) and inference in workflow?
- Publication of workflows in linked data format?
- Search queries for workflows of similar result, for similar tools, for similar data?
- Workflow, data or tool recommendation?